

Exercise 1

Lectures 1,2,3: Data Warehouses & Multidimensional Modeling

Scalable Data Engineering

1.1 Are the following statements true or false? Give reasons for your answer.

- a) Multidimensional modeling (MM) and data warehousing (DWH) are trivial problems.
- b) DWHs are set up and maintained by individual users.
- c) Virtual Integration combines all data sources into one central repository.
- d) A Data Mart is a context-based subset of a DWH.
- e) The Data Cube is the physical base for MOLAP and ROLAP.
- f) MOLAP is always better than ROLAP.
- g) The Star schema is a denormalized alternative of the Snowflake schema
- h) A Galaxy schema has exactly one fact table.
- i) A ROLAP schema is read-friendly.
- j) A ROLAP schema is update-friendly.

Please use <https://amcs.website> to prepare the answers for these questions. PIN: SDE22

1.2 Categorize the following schemes: Star schema, Snowflake schema, Galaxy schema, Snowstorm schema according to their structure and degree of normalization.

Degree of normalization	structure		

1.3 Transform the following table header into a Star schema.

- What could be interesting queries?
- Where are the redundancies in the given table?

movie title	kind	production year	season	episode	company	director	director status
string	category	number	number	number	string	string	category

1.4 Transform the following table header into a Snowflake schema.

- What could be represented in additional dimension tables?
- Hint: Use your results from 1.3

movie title	kind	production year	season	episode	company	director	director status
string	category	number	number	number	string	string	category

1.5 Compare the memory footprints of the Star and Snowflake schema.

- Use your solutions from 1.3 and 1.4
- Include a comparison for storing everything in one large table (cf. 1.3)
- Assumptions:
 - 1000 movies by 1000 directors and 1000 companies in fact table
 - No season or episode numbers
 - 64 bit integer and float
 - 8 bit strings with length 128
 - 7 kinds of movies
 - 3 states for director status